Draft Genome Assemblies of Xylose-Utilizing Candida tropicalis and Candida boidinii with Potential Application in Biochemical and Biofuel Production

Somani, Abhishek; Smith, Daniel; Hegarty, Matthew; Fernandez-Fuentes, Narcis; Ravella, Sreenivas; Gallagher, Joseph; Bryant, David

Published in:
Genome Announcements

DOI:
10.1128/genomeA.01594-17

Publication date:
2018

Citation for published version (APA):
Somani, A., Smith, D., Hegarty, M., Fernandez-Fuentes, N., Ravella, S., Gallagher, J., & Bryant, D. (2018). Draft Genome Assemblies of Xylose-Utilizing Candida tropicalis and Candida boidinii with Potential Application in Biochemical and Biofuel Production. *Genome Announcements*, *6*(7), [e01594-17]. https://doi.org/10.1128/genomeA.01594-17

AMERICAN SOCIETY FOR MICROBIOLOGY | genomeAnnouncements™

EUKARYOTES

Draft Genome Assemblies of Xylose-Utilizing *Candida tropicalis* and *Candida boidinii* with Potential Application in Biochemical and Biofuel Production

Abhishek Somani,[a] Daniel Smith,[a] Matthew Hegarty,[a] Narcis Fernandez-Fuentes,[a] Sreenivas R. Ravella,[a] Joe A. Gallagher,[a] David N. Bryant[a]

[a]Institute of Biological, Environmental & Rural Sciences, Aberystwyth University, Aberystwyth, United Kingdom

ABSTRACT Non-*albicans Candida* species are growing in prominence in industrial biotechnology due to their ability to utilize hemicellulose. Here, we present the draft genome sequences of an inhibitor-tolerant *Candida tropicalis* strain (Y6604) and *Candida boidinii* NCAIM Y01308^T.

Manufacturing higher-value commodities from hemicellulosic sugars (e.g., xylose) is crucial for environmental and bioeconomic sustainability of lignocellulosic biorefineries. *Candida tropicalis* has been widely investigated in the bioconversion of xylose into the higher-value sweetener xylitol and/or into bioethanol (1, 2), while the methylotroph *Candida boidinii* is well established for heterologous gene expression and enzyme/biochemical production (3, 4). In this study, we report the draft genome sequences of an environmentally derived inhibitor-tolerant *C. tropicalis* isolate (Y6604) and *C. boidinii* NCAIM Y01308^T (NCAIM, Budapest, Hungary). Strain Y6604 has high tolerance to lignocellulose-derived inhibitors (up to 3 g/liter furfural and 4 g/liter 5-hydroxymethylfurfural), and metabolically engineered variants have improved xylose-to-xylitol bioconversions in lignocellulosic hydrolysates (data not shown).

Genomic DNA (1 μg) (YeaStar; Zymo Research, USA) was extracted, and Illumina TruSeq libraries were size selected with AMPure beads for an average insert size of ~700 bp. Prior to sequencing on an Illumina HiSeq 2500 platform, paired-end reads were produced for both species, with additional mate pair reads for *C. boidinii* NCAIM Y01308^T. Overlapping paired-end sequence reads were merged using FLASH (5). The quality control suite FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) identified some initial 5′ base bias, low-quality 3′ bases, and Illumina adaptors, which were removed using Trimmomatic (6). NxTrim (7) was used for adaptor trimming of raw *C. boidinii* mate pair sequences. All sequences were error corrected and assembled using SOAPdenovo (8). Following genome masking, coding genes predicted by AUGUSTUS (9) and trained for *C. tropicalis* were submitted to a BLASTp search. Candidate hits (E value, $\leq 1 \times 10^{-10}$) were assigned names, Gene Ontology (GO) and Kyoto Encyclopaedia of Genes and Genomes (KEGG) annotations using Blast2GO (10).

A total of 51,470,621 paired-end reads for Y6604 gave an assembly length of 14,318,547 bp, with 296× coverage of 563 scaffolds (N_{50}, 51,027), and 10,177 contigs. Scaffolds with more than 50% "N" calls and <300 coding bases were removed, giving 533 scaffolds and 688 contigs. The combined *C. boidinii* assembly from both paired-end and mate pair reads had high coverage depths (259× and 523×, respectively) and a total length of 19,266,739 bases. Scaffolds with <1,000 coding bases and/or more than 50% N calls were removed, giving 79 scaffolds (N_{50}, 606,681) and 61 contigs. The overall GC content in Y6604 was 34%, while that in *C. boidinii* was lower (31%). *De novo* annotation with AUGUSTUS yielded 6,772 proteins in Y6604 (8,270 exons and 1,498

Received 6 January 2018 **Accepted** 9 January 2018 **Published** 15 February 2018

Citation Somani A, Smith D, Hegarty M, Fernandez-Fuentes N, Ravella SR, Gallagher JA, Bryant DN. 2018. Draft genome assemblies of xylose-utilizing *Candida tropicalis* and *Candida boidinii* with potential application in biochemical and biofuel production. Genome Announc 6:e01594-17. https://doi.org/10.1128/genomeA.01594-17.

Address correspondence to David N. Bryant, dgb@aber.ac.uk.

introns) and 6,067 proteins (6,951 exons and 884 introns) in *C. boidinii* NCAIM Y01308^T. Reciprocal best BLAST hits (maximum E value, 1×10^{-10}) for Y6604 proteins compared with the 6,254 proteins in the reference *C. tropicalis* (strain MYA-3404) assembly (11) identified 5,487 matching proteins, with 1,285 and 767 proteins unique to Y6604 and MYA-3404, respectively. A BUSCO (v3.0.1) (11) comparison between the 1,711 profiles within the order *Saccharomycetales* and the proteins predicted by AUGUSTUS suggested that the strain Y6604 and *C. boidinii* NCAIM Y01308^T gene sets were largely complete, with only 6% and 7% (96 and 112 genes, respectively) of the conserved orthologs, respectively, deemed missing.

Accession number(s). The *C. tropicalis* strain Y6604 and *C. boidinii* NCAIM Y01308^T whole-genome shotgun projects have been deposited in DDBJ/ENA/GenBank under the accession numbers PKKZ00000000 and PKKY00000000, respectively. The versions described in this paper are the first versions.

ACKNOWLEDGMENTS

This research was funded by Climate-KIC (Knowledge and Innovation Community) created by the EIT (European Institute of Innovation Technology) in ADMIT Bio-Succinnovate, with support from the ERDF/WEFO (BEACON). The Institute of Biological, Environmental & Rural Sciences (IBERS) received strategic funding from the BBSRC.

REFERENCES

1. Granström TB, Izumori K, Leisola M. 2007. A rare sugar xylitol. Part I: the biochemistry and biosynthesis of xylitol. Appl Microbiol Biotechnol 74:277–281. https://doi.org/10.1007/s00253-006-0761-3.
2. Mateo S, Puentes JG, Moya AJ, Sánchez S. 2015. Ethanol and xylitol production by fermentation of acid hydrolysate from olive pruning with *Candida tropicalis* NBRC 0618. Bioresour Technol 190:1–6. https://doi.org/10.1016/j.biortech.2015.04.045.
3. Osawa F, Fujii T, Nishida T, Tada N, Ohnishi T, Kobayashi O, Komeda T, Yoshida S. 2009. Efficient production of L-lactic acid by Crabtree-negative yeast *Candida boidinii*. Yeast 26:485–496. https://doi.org/10.1002/yea.1702.
4. Yurimoto H, Sakai Y. 2009. Methanol-inducible gene expression and heterologous protein production in the methylotrophic yeast *Candida boidinii*. Biotechnol Appl Biochem 53:85–92. https://doi.org/10.1042/BA20090030.
5. Magoč T, Salzberg SL. 2011. FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics 27:2957–2963. https://doi.org/10.1093/bioinformatics/btr507.
6. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170.
7. O'Connell J, Schulz-Trieglaff O, Carlson E, Hims MM, Gormley NA, Cox AJ. 2015. NxTrim: optimized trimming of Illumina mate pair reads. Bioinformatics 31:2035–2037. https://doi.org/10.1093/bioinformatics/btv057.
8. Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, Tang J, Wu G, Zhang H, Shi Y, Liu Y, Yu C, Wang B, Lu Y, Han C, Cheung DW, Yiu SM, Peng S, Xiaoqian Z, Liu G, Liao X, Li Y, Yang H, Wang J, Lam TW, Wang J. 2015. Erratum: SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler. Gigascience 4:30. https://doi.org/10.1186/s13742-015-0069-2.
9. Stanke M, Steinkamp R, Waack S, Morgenstern B. 2004. AUGUSTUS: a Web server for gene finding in eukaryotes. Nucleic Acids Res 32:W309–W312. https://doi.org/10.1093/nar/gkh379.
10. Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21:3674–3676. https://doi.org/10.1093/bioinformatics/bti610.
11. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31:3210–3212. https://doi.org/10.1093/bioinformatics/btv351.